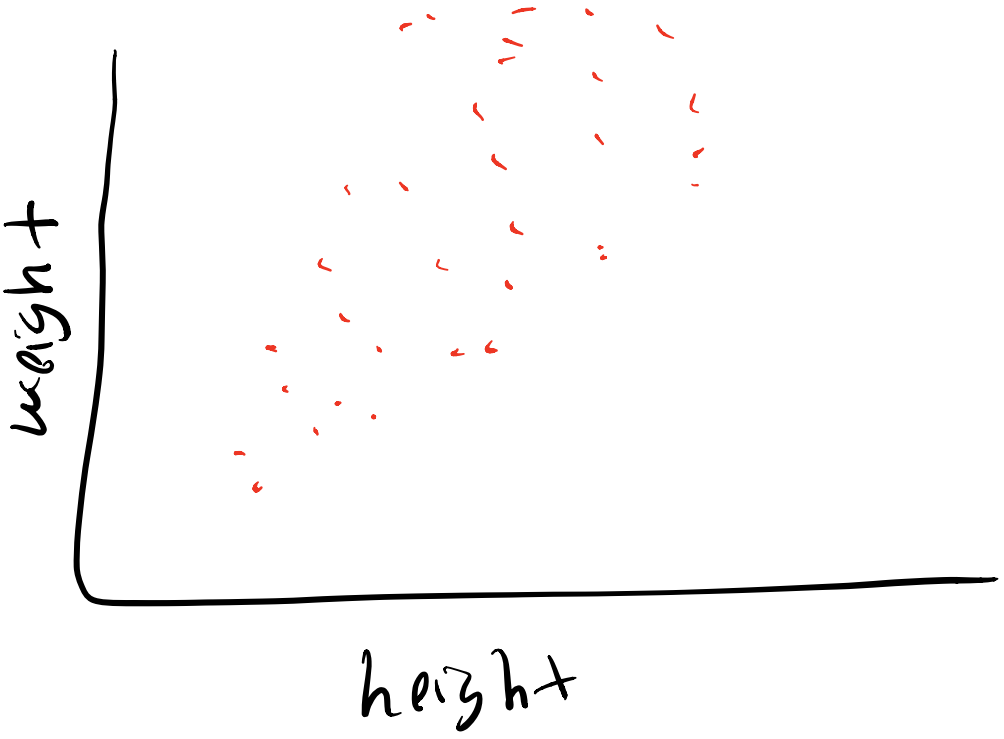
Correlation

Paired data can be plotted as a scatter plot



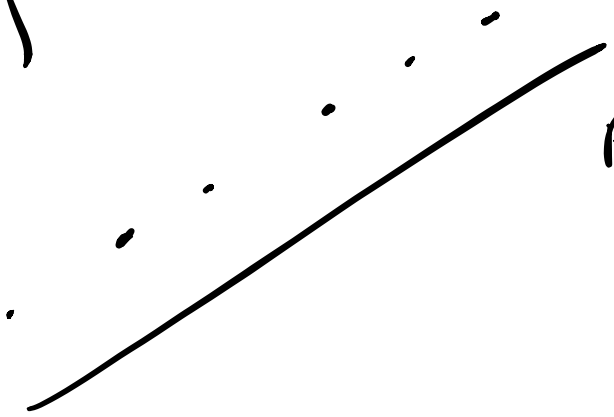
Each dot is a specific person

Correlation is "how close is this to a straight line"

Correlation between (r)

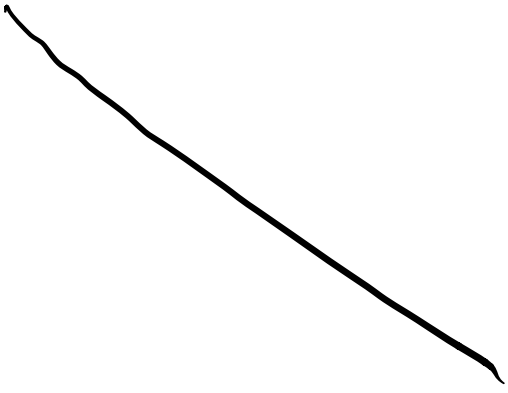
$[-1, 1]$

$r = 1$

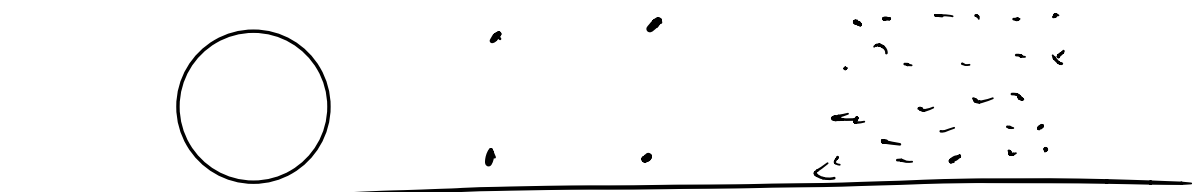


perfect straight line

$r = -1$ any neg slope perfect Straight line



$r = 0$, no linear relation



Computing cor (x, y)

Easiest way cor(x, y) in R.

x	y
x_1	y_1
x_2	y_2
⋮	⋮
x_n	y_n

1. Convert each column to SU (with their respective mean and sd)

2. multiply element wise to get new vector

$$z$$

$$su(x_1) \cdot su(y_1)$$

$$\vdots$$

$$su(x_n) \cdot su(y_n)$$

3. Take average of z.

Properties

adding a constant doesn't change r

multiplying by (+) constant doesn't change r.

any "linear" transform

$ax+b$ for $a>0$

won't change r

multiplying by (-) constant

flips sign of r.

$cor(x,y) = cor(y,x)$